

Contents

Contents	1
I Economics	5
1 Econometrics	7
1.1 Low and Meghir, 2017, JEP	7
1.2 Lewbel, 2016, JEL	8
1.2.1 Point Identification	8
1.3 Gentzkow-Kelley-Taddy, 17 WP	8
1.4 Seminar Notes	9
2 IO	11
2.1 Data and Techniques	11
2.1.1 Data Source	11
2.1.2 Techniques	11
2.2 Rey and Stiglitz, 1995, RAND	12
2.2.1 Benchmark Case	13
2.3 Production Function Estimation	14
2.4 Demand Function Estimation	15
2.4.1 Brief Literature Review	16
2.5 Price Dispersion	17
2.6 Multisided Platforms	18
2.7 Erdem-Keane, 96 MktSci	19
2.8 Gentzkov-Shapiro-Taddy, 17 WP	19
3 Game Theory	21
3.1 Important Tools	21
3.1.1 Solution Concept	21

3.2	Learning Theory	22
3.2.1	Important tools in learning	22
3.2.2	Overview of literature	23
3.2.3	Bikhchandani-Hirshleifer-Welch, 92 JPE	24
3.2.4	Kremer-Mansour-Perry, 14 JPE	24
3.3	Hard Information	25
3.3.1	Evidence Game	26
3.3.2	Beyer-Cohen-Lys-Walther, 10 JAcctE	26
3.3.3	Milgrom, 81 Bell	27
3.4	Green-Porter, 84 EMCA	28
3.5	Repeated Game	28
3.5.1	Imperfect Monitoring	28
3.6	Seminar Notes	29
4	Decision Theory	31
4.1	Dekel and Lipman, 2010, Annu Rev Econ	31
5	Matching Theory	33
5.1	Gale and Shapley, 1962, Ame.Math.Mon.	33
6	Search Theory	35
6.1	Shi, 09 EMCA	35
7	Macro	37
7.1	Kiyotaki, 2011	37
7.2	Costly State Verification	37
7.3	Townsend, 79 JET	38
7.4	Kiyotaki-Moore, 97 JPE	38
7.5	Gorton and Ordonez, 14 AER	39
7.6	Williamson, 87 QJE	39
7.7	Chari et.al, 14 AER	41
7.8	Euro Crisis	42
7.9	Seminar Notes	42
8	Methodology	45

<i>CONTENTS</i>	3
II Mathematics	47
9 General (mathematical) Skills	49
9.1 Log-Linearization	49
9.1.1 How to do Log-Linearization	49
9.1.2 log-linearization when there is expectation (An Example)	50
9.2 Supermodularity	51
10 Probability	53
10.1 Bayesian Statistics	53
10.1.1 Bayesian Updating	53
11 Numerical Methods	55
11.1 Numerical Maximization	55
12 Miscellany	57
Bibliography	59

Part I

Economics

Chapter 1

Econometrics

1.1 Low and Meghir, 2017, JEP

Low, H. and C. Meghir (2017, May). The Use of Structural Models in Econometrics. *Journal of Economic Perspectives* 31(2), 33–58

Defining a structural model:

A **fully specified model** make explicit assumptions about the economic actors' objectives and their economic environment and information set, as well as specifying which choices are being made within the model. They allow a complete solution to the individual's optimization problem as a function of current information set. Fully specified models are particularly useful in understanding mechanism of a policy, especially when we want to estimate some long-term effects of the policy.

A **partially specified model** relies on a sufficient statistic that summarizes choices not being modeled specifically. For example, assuming that the choices is only intratemporal instead of intertemporal.

Treatment effect models focus on identifying a specific causal effect of a policy while saying least about the theoretical environment. The pro is the cleanness of causality. The con is the limitation in exploiting the results outside. The identification of treatment model depends on assumptions that the experientment has not been compromised and there is no spillovers from the treatment units.

A combination of fully spfecied model and randomized experiments can enhance anaysis for both. Experimental evidence can be used either to validate a structural model, or to aid in the estimation process (in identification).

Solving structural models

This has been described in Adda and Cooper(2003) well. The general process is: 1) write down the bellmand function; 2) discrete the state space and decision space; 3) use value function iteration to solve the bellman function.

1.2 Lewbel, 2016, JEL

Lewbel, A. (2016). The Identification Zoo-Meanings of Identification in Econometrics. *Forthcoming on Journal of Economic Literature*

<https://www2.bc.edu/arthur-lewbel/ident-zoo-SL-Part1.pdf>

<https://www2.bc.edu/arthur-lewbel/ident-zoo-SL-Part2.pdf>

are links for a ppt on this paper by Lewbel.

There are two kinds of identification problems. 1. One is to identify the treatment effect, a typical example is the selection bias. The problem in these cases are that selection (determining who is treated or observed) and outcomes may be correlated. 2. Another is to identify the true coefficient in a linear regression when regressors are measured with error.

1.2.1 Point Identification

We start by assuming some information ϕ is knowable. A simple definition of point identification is that a parameter θ is point identified if, given the model, is uniquely determined from ϕ . Notice that this definition of point identification is recursive in some sense. To identify θ , we first need to assume some ϕ is knowable, which means ϕ itself is identified. This identification of ϕ can only be justified by further assumptions of DGP (Data Generating Process).

For example, for a model $Y = X\theta + e$, we assume that $E(e^2) \neq 0$ and $E(eX) = 0$, and suppose ϕ includes the second order of (Y, X) . Then we can conclude that ϕ is point identified, given by $E(XY)/E(X^2)$. Notice that the identification comes from the *assumptions* of model.

One common DGP is IID. Under this DGP, we can consistently identify the distribution of observation W . Another DGP is where each data point consists of a value of X chosen from its support, then we randomly draw Y conditional on X , which is independent from other draws conditional on this X . Under this DGP, we can consistently identify $F(Y|X)$. We can also use more complicated DGPs, for example, we generally assume only the second order moments are knowable in time series. One reason is this being sufficient for identification, another reason is higher order moments become unstable over time. Assumptions over GDP are always needed, even in experiment data, and which specific assumptions to take depend on the model.

1.3 Gentzkow-Kelley-Taddy, 17 WP

Gentzkow, M., M. Taddy, and B. T. Kelly (2017, February). Text as Data. pp. 1–53

An survey paper on how text analysis can be used in economic research. The main difficulty is the high-dimensionality of text.

Steps to reduce dimension:

1. Represent raw text \mathcal{D} as a numerical array C
 - (a) Divide \mathcal{D} into individual documents \mathcal{D}_i , the level is determined by the level of V . For example, daily attributes or monthly attributes.
 - (b) Feature Selection. Drop out punctuations, numbers, HTML tags, proper names and so on. Both very common and very rare words will be excluded. And words with the same stems will be reduced to one.
 - (c) Limit dependence between words using N-gram. The idea is to only consider phrase of length n .
2. Map C to predicted values \hat{V} of unknown outcomes V
3. Use \hat{V} in subsequent descriptive or causal analysis

Reading not finished. [Grimmer and Stewart \(2013\)](#) also talks this problem. worth reading

1.4 Seminar Notes

AI and Economics (<http://taddylab.com/slides/EconAI.pdf>)

This slides provides some clear intro on why and how AI (also deep learning) can be applied in economics. Many references provided in slides as well.

[Not understand much. Needs further reading.](#)

Chapter 2

IO

2.1 Data and Techniques

2.1.1 Data Source

Uber and Taxi

- New York Taxi and Limousine Commission Trip Data
 - http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml
 - Literature: Bian Bo (JMP)

2.1.2 Techniques

High Resolution Satellite Data:

- Pixel is as accurate as 0.5m, much finer than the Night Light data;
- Use machine learning to identify the roads, cars, buildings etc;
- Jonathan Hersh has a paper over this. **Poverty from Space: Using High Resolution Satellite Imagery for Estimating Economic Well-being and Geographic Targeting**

Paralelle Computation:

- A detailed cook manual on Paralelle computation of RBC model:
<https://www.kellogg.northwestern.edu/researchcomputing/docs/ParallelComputing/ParallelizationForRBC.pdf>
- Another source: https://bfi.uchicago.edu/sites/default/files/file_uploads/Lecture_Parallel_Programming_Chicago%20%282%29.pdf

Text as Data:

- Course Materials by Justin Grimmer
- Grimmer, J. and B. M. Stewart (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21(3), 267–297
- Gentzkow, M., M. Taddy, and B. T. Kelly (2017, February). Text as Data. pp. 1–53

2.2 Rey and Stiglitz, 1995, RAND

Rey, P. and J. Stiglitz (1995, October). The Role of Exclusive Territories in Producers' Competition. *The RAND Journal of Economics* 26(3), 431–451

For detailed proof of this paper, see Evernote.

Main result: vertical restrains can be used to reduce interband competition. Because exclusive territories alter the perceived demand curve, making each producer believe he faces a less elastic demand curve, inducing an increase in eqm price and producer's profits even in the absence of franchise fee. This result is different from traditional Chicago school results, which insist that exclusive territories will increase efficiency. This difference comes from market structure. Chicago schools investigate in full competition and full monopoly producer cases, while this paper looks at duopoly producer. In full competition and full monopoly case, the competition level has already been *preassumed*, while exclusive territories can reduce the competition level among producers in other cases.

The key for the result is the the following compound demand elastic:

$$\tilde{\varepsilon}(p^e) := m_1(p, p)\varepsilon_1(q, q) + m_2(p, p)\varepsilon_2(q, q)$$

where $q = q_1^r(p, p)$ (the response retail price), $m_1(p, p) = \partial \log q_1^r(p_1, p_2) / \partial \log p_1$ (the own elasticity of retail price to producer price), and $m_2(p, p) = \partial \log q_1^r(p_1, p_2) / \partial \log p_2$ (the cross elasticity of retail price to producer price), and $\varepsilon_1(q_1, q_2) = -\partial \log D^1(q_1, q_2) / q_1$ (the own elasticity of demand to retail price, positive), and $\varepsilon_2(q_1, q_2) = -\partial \log D^1(q_1, q_2) / q_2$ (the cross elasticity of demand to retail price, negative).

In the above equation, it is very reasonable to think $0 < m_1 < 1$ and $0 < m_2$. $m_1 > 0$ because the own elasticity of retail price to producer price is positive. $m_1 < 1$ means that retailer will absorb some increase in producers' price, which will be the case if demand elasticity becomes higher in high retail price. And $0 < m_2$ derives from the two products to be substitutes. Under these two conditions, combined with $\varepsilon_1 > 0$ and $\varepsilon_2 < 0$, then $\tilde{\varepsilon}(p^e) < \varepsilon_1$. Thus, under exclusive territories, the producers' perceived demand curve is less elastic. So the equilibrium price (both retail and producer) is higher even when no franchise fee applies.

Setting:

- two manufacturers produce imperfect substitutes at same marginal cost c
- retailers are perfect competition / or exclusive territory
- the final good demand depends on retail prices and is given by $D^i(q_1, q_2)$
- costs and demand functions are common knowledge, retailers observe all contracts signed by each producer
- producers only observe the quantity bought by retailers; they do not observe the quantities sold by retailers (i.e. fullline forcing is infeasible)
- producers serve many markets at no additional cost
- consumers have no search cost

Under these settings and information conditions, we can see it as a two stage game. In the first stage, producers simultaneously set wholesale price p_1 and p_2 . In the second stage, the retailer observe all wholesale price and decide the retail price simultaneously.

2.2.1 Benchmark Case

We use the following assumptions throughout the paper unless specified otherwise:

1. Let $\pi(p_i, q_1, q_2) := (q_i - p_i)D^i(q_1, q_2)$ denote the retail profit for product i ; assume it to be twice differentiable wrt each argument, and is single peaked wrt q_i . The reaction function $q_i^a(p_i, q_j)$ is thus continuously differentiable and characterized by FOC.
2. Products are substitutes: $\partial D^i / \partial q_i \leq 0$ and $\partial D^i / \partial q_j \geq 0$
3. Demand functions are symmetric: $\forall p_1, p_2 \in \mathbb{R}_+, D^1(p_1, p_2) = D^2(p_2, p_1)$

Think of the benchmark case that no vertical restriction so the retailers are perfect competitive, and the producer monopolizes. In this case, the game is just a one step optimal pricing problem. The producer chooses an optimal retail price.

Useful trick:

Throughout this paper, we can write the symmetric eqm conditions in the following form:

$$(p^c - c)/p^c = 1/\varepsilon(p^c, p^c)$$

where p^c is the symmetric eqm price, and ε is some kind of elasticity.

In a symmetric eqm, the FOC of each producer gives $(p^c - c)/p^c = 1/\varepsilon_1(p^c, p^c)$ where $\varepsilon_1 = -\partial \log D^1(q_1, q_2)/\partial \log q_1$, i.e. the self elasticity of demand.

In the simplest benchmark case, the two factories are integrated, leading us to: $(p^c - c)/p^c = 1/E(q^m)$, $E(q) := \varepsilon_1(q, q) + \varepsilon_2(q, q)$, where $\varepsilon_2 = -\partial \log D^1(q_1, q_2)/\partial \log q_2$ (the cross demand elasticity).

In the exclusive territory case, the symmetric eqm satisfies $(p^e - c)/p^e = 1/\tilde{\varepsilon}(p^e)$, where

$$\tilde{\varepsilon}(p^e) := m_1(p, p)\varepsilon_1(q, q) + m_2(p, p)\varepsilon_2(q, q)$$

where $q = q_1^r(p, p)$ (the response retail price), $m_1(p, p) = \partial \log q_1^r(p_1, p_2)/\partial \log p_1$ (the own elasticity of retail price to producer price), and $m_2(p, p) = \partial \log q_1^r(p_1, p_2)/\partial \log p_2$ (the cross elasticity of retail price to producer price).

2.3 Production Function Estimation

There are several difficulties in estimating production functions which will cause endogeneity problem. To name some:

- Simultaneity. For a production function like $y_{it} = \beta_0 + \beta_1 k_{it} + \beta_2 l_{it} + \omega_{it} + \varepsilon_{it}$, we suppose that ω_{it} is observed by firm but not by econometricians, and ε_{it} is observed neither by firm nor econometricians. This could induce endogeneity when directly using OLS. Suppose ω_{it} has *positive serial correlation*, when ω_{it} is high, firm will choose higher input because it is very possible that ω in next period will be high as well. This induces a positive bias for β_k .
- Exit selection bias. When constructing a balanced panel, then we only observe those firms with high ω_{it} . [Why we have to construct a balanced panel?](#)

There are several approaches in solving this problem.

- Classical Approach.
 - Use input price as IV. The problem is input price is generally the same across firms. When we do observe different input prices, which generally indicates market power in input markets. Thus those facing higher ω_i will tend to produce more, inducing their input price to increase, which has a negative bias on β_k .
 - * (My comment) When using input price as IV, we actually assume they are independent of productivities. This may not be the case. If productivity has some correlation across the firms, then it is possible that when productivity is high, the input price in the market will also be high.
 - Use fixed effect. We need to assume that a firm faces same ω_{it} across time, which is not very practical.

- Control Function for ω_{it} . This approach first raised by Olley and Pakes (1996) and developed by Levinsohn and Petrin (2003) and Akerberg et al. (2015). We name them OP, LP, ACF here after. The key idea is to use a function of what we know (for example, k_t and i_{t-1} in OP) to proxy ω , the existence of such function is guaranteed by the solution property of the firm optimization problem. And the specific form of this function is estimated non-parametrically.
 - OP uses a two step estimation process. First step, use k_t and i_t non-parametrically estimate ω_t , plug this into original production function then use OLS. Notice the unobserved ω_t has been controlled by k_t and i_t , then OLS will give consistent estimator for l_t but not k_t . In second step, we first write $\omega_{it} = E[\omega_{it}|\omega_{it-1}] + \xi_{it}$, notice that ξ_{it} is independent of any information set at t-1. Thus we get the moment condition $E[\xi_{it}k_{it}] = 0$, then we can use method of simulating moments to estimate β_k .
 - * The key assumption in first step is, current period investment i_t can be represented by a function of $f(k_t, \omega_t)$. If f is invertible in ω_t (for example, f being monotone in ω), then we can write ω as a function of (i_t, k_t) .
 - LP is very similar to OP, just that use intermediate input m_{it} instead of investment i_{it} in proxy ω_{it} . Because OP approach requires the policy function of i_{it} be monotonicity in ω_{it} , which is unsatisfied when $i_{it} = 0$. But data shows that in many observations $i = 0$. Using intermediate input can solve this.
 - ACF criticizing is that β_l cannot be identified in the first step. In LP we assume that $m_{it} = f_t(k_{it}, \omega_{it})$. But we can reasonably argue by the same reason that $l_{it} = h_t(k_{it}, \omega_{it})$. Then since $f(\cdot)$ is invertible, then we can write l_t as some deterministic function of k_{it} and m_{it} . This implies we cannot identify β_l in the first step. The solution of ACF is to identify both β_k and β_l simultaneously at step 2, which is feasible since we now have two moment conditions.
- Gandhi et al. (2017) Approach (GNR). GNR criticizes control function approach is nonparametrically not identified in the presence of flexible inputs. They raise a new approach using FOC conditions.

2.4 Demand Function Estimation

Reference: Shum (2016) CH 1 is very good for motivation and techniques.

Motivation for demand estimation

- Firstly, it is important per se, we want to understand consumer behaviour. And we want to *predict the results of price change*, and *introduction of new products*. Such counterfactual predictions can only be built upon a structural model.

- Secondly, estimation of demand function helps us better understand market structure. Markup itself is rarely observed (with some exceptions, like the electronic industry), but we can generally observe demand elasticity. Recall that $\frac{p-MC}{p} = 1/\varepsilon$ (see 2.2). Estimation of ε will help us know the markup of firm, thus infer market structure.

Challenges in demand function estimation:

- Too many parameters. Think of a market with 100 commodities and we want to estimate the cross elasticities. The number of parameters will be 10,000. Obviously it is impossible to have enough data to estimate so many parameters.
 - Solved by nested logit model.
- Unobserved quality. This will causes the endogeneity of price. And because the unobservable quality affects the selling in a highly nonlinear way, so IV is not very applicable in this case.
 - Solved by BLP style estimation.

2.4.1 Brief Literature Review

There are two branches of demand function estimation. The early literature features differentiated goods (continuous choices). Classical examples include Hausman et al. (1994) introducing a nested approach to reduce the dimension of parameters.

Recent literature focus much more on discrete choice model. The seminal papers are Berry (1994) and Berry et al. (1995). The address of these two papers is to overcome the **endogenous price problem due to unobserved quality**. Suppose we use a characteristic approach, where consumer utility and thus their purchase decision is decided by characteristics of products. Some of the characteristics are observable to consumer (and firms) but not to econometricians, for example the advertisement. For higher unobserved quality, the equilibrium price will be higher, leading to a downward bias of price elasticity.

Berry (1994) comes with a method to overcome this endogeneity problem with only *market level data*.

- The basic idea is a IV based two step estimation.
- First step: denote δ_j to be the ‘mean utility’ for brand j . Notice this includes unobserved quality. Under some conditions, δ_j can be *inversed* by market shares (s_1, \dots, s_J) , and thus we can get $\hat{\delta}_j$ by observed market shares. [This inversion step is the key for this paper, and many following papers treating endogenous quality problem.](#)
- Second step: By definition, $\delta_j = X_j\beta - \alpha p + \xi_j$. Thus we can get $\hat{\xi}_j$ if we get $\hat{\delta}_j$. Suppose $E[\xi Z] = 0$, then we can use Z as IV to estimate consistently.

Berry et al. (1995) (BLP henceforce) is more comprehensive with the same idea. They introduce random-coefficients logit model, which allows each individual to have different taste coefficients. Thus we can’t

directly use the inverse technique. The way is to use a Simulation Method of Moments (inner loop and outer loop). First guess a δ , match mkt share, then run a IV regression, then guess different δ and get an outer loop. They prove a contraction mapping.

Quan, T. W. and K. R. Williams (2017). Product Variety, Across-Market Demand Heterogeneity, and the Value of Online Retail. *SSRN Electronic Journal*, 1–74

This paper addresses the local heterogeneity of consumer demand, and concludes that usual BLP method will overestimate the welfare increase of e-commerce because local stores have satisfied most local demands.

Challenges:

- Huge varieties of goods in fine-grid locations make scarce data in each category. Two possible solutions for this: 1) aggregating over locations; 2) just ignore the zero sales. Aggregating is unsatisfactory since the heterogeneity of local preference is our study focus. Ignoring will cause selection bias and cause overestimation of local heterogeneity.
- Mean utility level cannot be recovered solely by observable variables. They raise specific Converging in Distribution theorem to tackle this.

2.5 Price Dispersion

Baye, M. R., B. De los Santos, and M. R. Wildenbeest (2013). The Evolution of Product Search. *The Journal of Law, Economics Policy* 9(2), 1–21

- Price comparing sites becoming less popular, more consumers choose to directly go to like Amazon and eBay. These websites also have ‘marketplace’.
- Many searching behaviour are not observed. Non-texture data (e.g. in the menu, directories).
- Search time across platforms can hardly be compared. Because maybe it’s enough to search only once in Amazon, but many times in eBay.

Random Ideas

- Collusion when there is consumer searching cost. Intuitively, collusion should be easier to sustain with [\[no idea\]](#)
- Awaya-Krishna (2016)

2.6 Multisided Platforms

Schmalensee, D. S. E. and Richard (2013, October). The Antitrust Analysis of Multi-Sided Platform Businesses. *manuscript*, 1–45

- Price structure matters. Reduce charge by one side and increase charge by another side will affect the transaction.
- Transaction cost is necessarily high for multisided platform.
- Two kind of externalities. Usage externality and membership externality.
- Pricing is complicated. Per-transaction fee, or membership-fee. [This looks similar to conventional IO model.](#)
 - In per-transaction fee case. Price is directly proportional to the elasticity, while in conventional models price is inversely related to elasticity.

$$\frac{p_1 + p_2 - c_1 - c_2}{c_1 + c_2} = \frac{1}{e_1 + e_2}$$

$$\frac{p_1}{e_1} = \frac{p_2}{e_2}$$

Notice the second equation is unconventional.

- In membership-fee case.
- Both models can lead to a highly skewed price structure.
- Competition Structure.
 - Product differentiation
 - Multi-home.
 - Asymmetric Competition. A n-sided platform may face competition from 1) one-sided firms; 2) a multisided platform competing in some but not all sides; 3) a multisided platform that has additional sides

Random Ideas

- Multisided platform as a commitment device changes the game structure. Like in the Truck-GasStation case raised in the book, the fleet-card company provides commitment power to the card company, no renegotiation problem.

2.7 Erdem-Keane, 96 MktSci

Erdem, T. and M. P. Keane (1996). Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science* 15(1), 1–20

Erdem and Keane (1996) provides an example on how to estimate demand when quality is uncertain and consumer learns by Bayesian rule.

Motivation:

- Avoid ad hoc assumptions on how consumers utilizing experience information
- Estimate structural models to do policy analysis

Features:

- Estimate Dynamic Programming. Consumers make decision based on their perceptions, but econometricians only observe the consumption. So have to integrate over the distribution of perception errors $v_{ij}(t)$, which is a Markov process. The integration part will be very complicated.
 - Provides a way to estimate Dynamic Programming model with simulation. Also see Keane and Wolpin (1994).
- Experience Shocks. This indicates that even after purchasing once, consumer is not certain about the quality of the good (or the preference of himself). This is mainly motivated by empirical fact: consumption behavior generally do not converge.

Empirical Findings:

- Bayesian learning very well fit the data.

2.8 Gentzkow-Shapiro-Taddy, 17 WP

Gentzkow, M., J. Shapiro, and M. Taddy (2017, May). Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech

- Methods for analyzing high-dimensional choices decisions
 - also see Gentzkow, M., M. Taddy, and B. T. Kelly (2017, February). Text as Data. pp. 1–53
- Index for measuring polarization

This is a methodology paper introducing how to analyze high-dimensional choices decisions. They show that traditional logit model has severe finite sample bias. They also raise a index for measuring polarization.

[TO BE ADDED]

Chapter 3

Game Theory

3.1 Important Tools

Theorem 3.1.1 (Revelation Principle, Myerson, 1981). *Suppose that ψ was a Bayes Nash equilibrium of the indirect mechanism Γ . Then there exists a direct mechanism that is payoff-equivalent and where truthful revelation is an equilibrium.*

Proof: [to be added]

□

Comment:

- Revelation Principle is defined over some specific solution concept. Here it is Bayes Nash Eqm.
- Revelation principle is not always satisfied. The stronger solution concept we use, the more likely it is not satisfied.

3.1.1 Solution Concept

There are many solution concepts:

- Nash eqm
- Subgame Perfect eqm
- Perfect eqm
- ε -Perfect eqm
- Proper eqm
- Trembling-Hand eqm
- Bayesian Nash eqm

- (Weak) Perfect Bayesian eqm
- Sequential eqm
- Forward looking eqm
- Self-Confirming eqm

Definition 3.1.1 (Self-confirming Equilibrium, Fudenberg and Levine (1993)). Profile σ is a *self-confirming equilibrium* if for each player i , $s_i \in \text{support}(\sigma_i)$ there exists belief μ_i such that:

1. s_i maximizes $u_i(\cdot, \mu_i)$, and
2. $\mu_i[\{\pi_{-i} | \pi_j(h_j) = \hat{\pi}_j(h_j | \sigma_j)\}] = 1$, for all $j \neq i$ and $h_j \in \bar{H}(s_i, \sigma_{-i})$, where \bar{H} is the set of info sets that are reached with positive probability

Comment:

- This definition mainly requires that the belief μ_i is consistent only at information sets that are reached with positive probability when player i plays s_i .
- This is most obvious when contrast with Nash Eqm, which is almost the same as this definition, except for the consistency requirement is for all $h_j \in H_{-i}$. The consistency is required for all information sets.
- Self-confirming is a weaker solution concept than Nash Equilibrium. It uses the idea that players can only learn through what they observe. For those outcomes that they cannot observe, there is no particular reason for their beliefs to coincide. Players can never learn something they lack information on.

Example 3.1.1.

3.2 Learning Theory

3.2.1 Important tools in learning

Definition 3.2.1 (Monotone Likelihood Ratio Property). For all $l < S$,

$$\frac{p_{q,l}}{p_{q+1,l}} \geq \frac{p_{q,l+1}}{p_{q+1,l+1}} \text{ for all } q < R$$

with strictly inequality for at least one q .

This is first raised by Milgrom (1981), which ensures that the conditional expectation of each individual increases in his signal realization.

Definition 3.2.2 (Martingale). A *martingale* is a pair (X_n, \mathcal{F}_N) , where $\{\mathcal{F}_n\}$ is a filtration and X_n an integrable (i.e. $E|X| < \infty$) stochastic process adapted to this filtration, such that

$$E[X_{n+1}|\mathcal{F}_n] = X_n, \text{ a.s. } \forall n$$

Comment:

- Notice that Martingale process is much more restricted than Markov process, not only in the realization of expectation value. Markov process requires that

$$Pr(X_{t+1} \in A) | X_t = x_0, X_{t-1} = x_{t-1}, \dots, X_{t-k} = x_{t-k} = Pr(X_{t+1} \in A) | X_t = x_0$$

for all $A \in \mathcal{X}$, all $x_0, x_1, \dots, x_k \in \mathbb{X}$, and all $0 \leq k \leq t$.

- This actually restricts all stochastic properties, including expectations, variations and others.
- Also, Martingale process require conditional on \mathcal{F}_t , but Markov process only requires x_t .

Theorem 3.2.1 (Martingale Convergence Theorem).

3.2.2 Overview of literature

The equilibrium concept used in literature is Bayesian Nash Eqm. Thus we put no restriction on off-eqm path beliefs. Nageeb said something about this, but I don't understand why this is important.

Bikhchandani et al. (1992) (BHW here after) is one of the original papers for learning in games. The main point is to explain the easily formulated but fragile information cascade when only actions are observed. The idea is, if only actions are observed, then after some threshold, private information will not affect action. And this private will NOT join the information pool, thus all following individual's private information will not contribute to the action. But since very little information is needed to form a cascade, this cascade can also be easily reversed when some shocks happen.

Smith and Sørensen (2000) is a generalization of BHW paper. The generalization lies in these aspects:

- Heterogeneity in quality of information (possibly unbounded private information)
- Heterogeneity of preference (noise by crazy types)
- Information cascade is NOT a generic feature; but 'herding' is.
- Necessary and sufficient condition for *complete learning*.

This paper also provides a general approach on how to handle learning behaviour.

3.2.3 Bikhchandani-Hirshleifer-Welch, 92 JPE

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy* 100(5), 992–1026

BHW is the seminal paper of social learning, which explains imitating in social phenomena without any ad hoc assumptions. By ad hoc assumptions, I means following:

- sanctions on deviants
- positive payoff externalities
- conformity preference

These assumptions all have the unsatisfying property that puts the cart before the horse.

TO BE ADDED

3.2.4 Kremer-Mansour-Perry, 14 JPE

Kremer, I., Y. Mansour, and M. Perry (2014). Implementing the “Wisdom of the Crowd”. *Journal of Political Economy* 122(5), 988–1012

Different from previous learning papers, KMP focuses on the optimal information disclosure of planner. They find that it is not always optimal for planner to disclose full information, because it will discourage agents’ incentive to explore and exploration has positive externalities.

Example: Online review website like TripAdvisor, Yelp; Google Map

Illustrative example:

Say R_1 is uniformly distributed in $[-1, 5]$, and R_2 is uniformly distributed in $[-5, -1]$. Then agent-1 for sure will choose a_1 since its ex-ante expected payoff is higher. Planner cannot force agent-1 to choose a_2 because of IC condition.

If realized R_1 is lower than 0, then for sure agent-2 will choose a_2 . But planner can do better by recommending agent-2 to play a_2 whenever $R_1 \leq 1$. Agent-2 will follow this suggestion because the $E[R_1]$ is 0 conditional on planner not disclosing.

If unfortunately the realized R_1 is higher than 1, then there is no way to make agent-2 to choose a_2 , but it is still possible to ask agent-3 to explore if R_1 is not too high. Because when player-3 is recommended to play a_2 , she cannot distinguish between: 1) planner knows both realization and $R_2 > R_1$; or 2) planner wants her to explore. The second situation is in the same shoes as agent-2, but because of the possibility of the first situation, agent-3 will admit a higher upper threshold for exploring.

□

Main Results: (The proof of this paper is not difficult, and worth learning)

- The optimal policy is threshold policy. For each individual, there is an interval of realization r_1 such that he is the first agent to explore.
- There is an upper bound of exploration time, which is independent of agents number T and realization of R_1, R_2 . This indicates result will converge to first best when T is large enough.

Comments:

- Generalize action to n will not affect the structure of optimal policy. While continuum of action will ask for new setting. See referee report for more.
- We can also think the realization of R_1 and R_2 is **correlated**. The result of this paper will not hold in this case. Think of the extreme case that correlation between R_1 and R_2 is -1 . Then for any $R_1 > 0$, the planner will never explore a_2 . For any $R_1 < 0$, the planner will recommend a_2 for all agents except for first one, and they will accept this recommendation. The correlation problem also exists when the realization of R_i is not a value but a distribution, in the sense of correlation in mean value. In another direction, if the realization is positively correlated, then the exploration cost will be lower. And optimal policy should have lower upper threshold for each agent.

I think this extension important, because we have correlations in many real world applications. Take TripAdvisor as an example. If we already have reviews on one Marriott hotel in Paris, then it's not very useful to explore another Marriott at Toulouse! Intuitively, planner should utilize such correlation, and put more weight on exploration of independent realizations. The optimal policy could be interesting.

- I also think of how this can be combined with optimal persuasion. If planner and agents have discrepancy in interest, then persuasion is needed.

3.3 Hard Information

Unraveling:

Unraveling is first discussed by Grossman (1981) and Milgrom (1981). They show that if investor knows the type of firm's information, such information is perfectly verifiable at no cost, and the disclosure of such information is at no cost. Then exist a sequential equilibrium, where the firm will disclose all information, good or bad. The intuition is, since investors know the type (distribution) of information firm has, the nondisclosure indicates a worst information. If the investor is so suspicious, firm will disclose those not so good information.

Dye (1985) finds that such unraveling may not always happen. There are many possible reasons, basically they are violations of above three assumptions.

Quigley and Walther (2017) discusses unraveling in existence of outside information. They show that such information beyond control of insiders will *reverse* unraveling result. The highest quality insiders remain

quiet, and only the mediocre insiders disclose information. Notice that sending message is costly in this model. The intuition for this result is, when type is high enough, outside information is enough to persuade the outsider. My guess is, this holds because the outsider decision is not fully divisible.

Mechanism Design with Hard Information:

Glazer and Rubinstein (2004) discusses the optimal decision rule when listener can check only one value. [\[Not yet read\]](#)

Green and Laffont (1986) discusses the direct mechanism under partially verifiable information. They show that the *Nested Range Condition* is necessary and sufficient for all socially-implementable mechanisms to be truthfully-implementable mechanisms. The condition requires that if A can pretend to be B, B can pretend to be C, then A can pretend to be C. This condition is used in many other papers like Hart et al. (2017), and monotonicity condition in Grossman (1981) is also included. But notice that their paper does not consider an explicit checking process, so the result cannot apply to Ben-Porath et al. (2014), where a perfect but costly checking technique exists.

3.3.1 Evidence Game

[\[Below is based on the talk by Xiao\]](#) The paper by Hart et al. (2017) provides a general framework for evidence games. It shows that under conditions:

- A1. $\theta \in m(\theta) \subseteq \Theta$. It basically restricts the message space to be the type space, and every can truthfully report his own type.
- A2. If $\theta' \in m(\theta'')$ and $\theta \in m(\theta')$, then $\theta \in m(\theta'')$. It means the masquerade is transitive, which is quite restrictive condition.

With above two assumptions, and *truth-leaning* refinement of equilibriums, they show that giving receiver commitment power will make him no strict better off.

Many more basic models like Dye (1985) and Milgrom (1981) can be encompassed by this framework. [\[Worth trying to put them in.\]](#)

3.3.2 Beyer-Cohen-Lys-Walther, 10 JAcctE

Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther (2010, December). The financial reporting environment Review of the recent literature. *Journal of Accounting and Economics* 50(2-3), 296–343

This paper surveys literature of **voluntary information disclosure**, which is highly correlated with *hard information* literature we discuss in reading group. I mainly read the section 3 (voluntary disclosure) of this paper.

We have unraveling result in Milgrom (1981) and Grossman (1981). Unraveling means the agent will disclose all their private information. This result relies on the following conditions:

1. disclosures are costless;
2. investors know that firms DO have private info;
3. all investors interpret the firm's disclosure in the same way, and firms know how investors will interpret that disclosure;
4. agents want to maximize their firm's share prices (no moral hazard problem);
5. agents can credibly disclose their private info;
6. firms cannot commit ex-ante to a specific disclosure policy.

Violation of any conditions will result in partial nondisclosure of private information.

1. *Costly disclosure.* If disclosure is costly to a firm, then it will choose not to disclose even when signal is not that bad.
2. *Probabilistic information endowment.* If firms get private information in probability, then they have incentive to not report some bad results. Because investors cannot distinguish them from those without private information.
3. *Uncertain investor response.*
4. *Uncertain disclosure incentives.* In this case, there will be moral hazard problem.
5. *Non-verifiable disclosure.* There are many important models in this category. For example, Cheap Talk model and Costly State Verification models. In cheap talk, exists "babbling eqm" since signal can't be verified. In Costly State Verification, firms have some rent since it's not wise to investors to check at every case.
6. *Ex-ante commitment.*

3.3.3 Milgrom, 81 Bell

Milgrom, P. R. (1981, October). Good News and Bad News: Representation Theorems and Applications. *The Bell Journal of Economics* 12(2), 380–391

Milgrom (1981) introduces a way to measure how good a message is. The difficulty here is that we are comparing two conditional distributions, not two values, thus something similar to FOSD is needed.

Definition 3.3.1. A signal is *more favorable than* another signal if the posterior belief $G(\cdot|x)$ FOSD $G(\cdot|y)$.

Proposition 3.3.1. x is more favorable than y iff for every $\theta_1 > \theta_2$,

$$\frac{f(x|\theta_1)}{f(x|\theta_2)} > \frac{f(y|\theta_1)}{f(y|\theta_2)} \quad (3.1)$$

Intuitively, x is more favorable iff x is more likely to appear when the state is high.

Definition 3.3.2 (Strict Monotone Likelihood Ratio Property). The family of densities $\{f(\cdot|\theta)\}$ have the monotone likelihood ratio property if for every $x > y$ and $\theta_1 > \theta_2$, (3.1) is satisfied.

Densities with MLRP has the good property that any two signals are comparable.

3.4 Green-Porter, 84 EMCA

Green, E. J. and R. H. Porter (1984). Noncooperative Collusion under Imperfect Price Information. *Econometrica* 52(1), 87–100

Motivation:

- In perfect monitoring models, the price is constant over time. But empirically, we observe price war frequently.
- Using imperfect monitoring, the price war is an intrinsic property of the model.
- The idea is: based on imperfect monitoring. The firms cannot perfectly distinguish others' deviation with a price shock.
- So there will be an price threshold. Above this threshold, they take the collusion action. Below this, they will exert retaliation for some period, then back to collusion.

Model:

-

3.5 Repeated Game

This section highly depends on Nageeb's notes and Mailath-Samuelson book.

3.5.1 Imperfect Monitoring

- Nageeb's notes

Motivation for imperfect monitoring:

- In practice, monitoring is often imperfect (Green and Porter (1984))
- [NOT UNDERSTAND](#). Helps to understand punishments at off-path histories

With imperfect monitoring, ‘forgiveness’ will often make the equilibrium better while keeping IC satisfied.

The difficulty in describing imperfect monitoring game with automata is, we can only observe ex-post history, how can we use automata graph to describe this ex-post history? The answer is we can think the equilibrium (and strategy) ex-ante. Because everything is linear, so it actually does not really matter what ex post realization is.

3.6 Seminar Notes

George Mailath (WP 17): “The Wisdom of The Confused Crowd: Model-Based Inference”

Main idea:

- If players are non-Bayesian, how do they update? How they exchange information in the information mkt?
- Agents do ‘model-based’ inference. Basically this is a two step structure.
- Agents have a model which addresses on some dimension of the information, and when they receive an information, they will only first focuses on the dimensions that their model addresses.
- Then they exchange the posterior belief with other agents. Since other agents have different model addressing different aspects, their posterior belief reveals some information of other dimension signal.
- Since the signals in different dimensions are correlated, one agent may use other’s posterior belief to infer partial information of signals in the dimensions that he is interested in.
- Thus, although they have different models addressing different dimensions, they can still exchange in the info market.

Simon Board, Moritz Meyer-ter-Vehn (WP 17): An Inspection Model of Learning in Networks

Main idea:

- Note of Takl by Board
- Two concerns in network learning
 - I get awaked at a constant rate, and may want to try
 - I see my neighbor trying, and want to try
- Learning in small network is very difficult to numerically computing.
 - Self-reinforcement structure

- Triangle correlation
- \Rightarrow can characterize, but cannot compute
- In large random network, such concern disappears
 - Because in large random network, the prob of each two points are connected is the same. And with large enough network, the prob of a small local structure is second order small.
 - Can write the graph as a tree in this case
- (?) Learning rate increases exponentially with the number of neighbors.

Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2017). Revealed price preference: Theory and stochastic testing

[Important paper, worth reading.](#)

This paper raises a new representation theorem over preference and utility. Traditionally we use preferences over consumption bundles x , and Afriat theorem guarantees a well-behaving utility representation if GARP is satisfied for preference. This paper raises a representation theorem for preference over price bundles.

More specifically, GARP says if we have $p^{t'} \cdot x^t < p^{t'} \cdot x^{t'}$, then we conclude $x^{t'}$ is revealed preferred to x^t , because x^t is affordable when the consumer choose $x^{t'}$. To test this approach, econometricians have to observe the consumption bundles of *all* goods, which is infeasible in many cases.

This paper raises another way to view data.

- If $p^{t'} \cdot x^t < p^t \cdot x^t$, then we say $p^{t'}$ is revealed preferred to p^t . Because the consumer has money left for other goods when price is $p^{t'}$. In this way, we can construct the preference over prices.
- This paper raises GAAP corresponding to GARP, and augmented utility function (seems integrating expenditure directly in the utility) w.r.t. GAAP.
- This representation can be empirically tested without observing all goods being purchased. And solves the endogenous expenditure level problem, which McFadden-Richter estimation approach assumes.
- Ron points out that this representation model cannot account for budget change.

Cooperation Game:

Nageeb mentions in class that one reason for cooperation game to be unpopular is the difficulty in treating informations. How to treat to coalition when someone has information while others don't?

Chapter 4

Decision Theory

4.1 Dekel and Lipman, 2010, Annu Rev Econ

Dekel, E. and B. L. Lipman (2010). How (not) to do decision theory. *Annual Review of Economics* 2(1), 257–282

What we can learn from decision theory

1) to know whether our intuition is correct; 2) to flesh out initial intuition to get additional or better prediction (e.g. some additional observable prediction); 3) to help us understand of why and how mechanism works

Story of the model

Why the story of the model is important even if it is not realistic? Why we can't just choose the model by the 'prediction-rejection' procedure?

1) The story makes us believe the predictions more; 2) Having a nice intuition helps us to utilize the model and to expand the model; 3) And a model can never be realistic, the important thing is whether it captures something important.

Chapter 5

Matching Theory

5.1 Gale and Shapley, 1962, Ame.Math.Mon.

Gale, D. and L. S. Shapley (1962, January). College Admissions and the Stability of Marriage. *The American Mathematical Monthly* 69(1), 9–15

This paper introduces Deferred Acceptance (DA) algorithms in matching, and shows that it is stable and optimal.

Definition 5.1.1. An assignment is called **unstable** if there are two applicants α and β who are assigned to universities A and B, although β prefers A to B and A prefers β to α .

Definition 5.1.2. A stable assignment is called **optimal** if *every* applicant is at least as well off under it as under any other stable assignments.

Comment: it looks at first that optimal assignments may not always exist. But as we will show constructively, it does.

Definition 5.1.3. Deferred Acceptance algorithm works as following. First let each boy propose to his favorite girl. Each girl who receives more than one proposal rejects all but her favorite boy. Yet, she doesn't accept him, but keeps him in a string to allow for the better may come later. Then in the second round, each rejected boy propose their second favorite girl, and each girl picks her favorite among the new proposers and the one in the string, and put him in her string. This process continues, until each girl has one boy in the string (this will happen after finite rounds), then each girl picks the boy in her string.

We will show the assignment by above algorithm is stable and optimal.

Chapter 6

Search Theory

6.1 Shi, 09 EMCA

Chapter 7

Macro

7.1 Kiyotaki, 2011

Kiyotaki, N. (2011). **A mechanism design approach to financial frictions.** *manuscript*

This paper mainly discusses the micro structures which may induce financial friction (inefficiency). These structures include private information of income, private technology for storage, limited commitment and their combinations.

7.2 Costly State Verification

Financial frictions is an important yet not well explained phenomena. Generally financial frictions means some possible financial frictions not happening, resulting in fluctuations in real economy. There are several micro foundations for financial frictions, we learned some in Ruilin's course:

- costly state verification
- collateral constraints
- moral hazard
- asymmetric information

The main framework for costly state verification is following. The borrower has some private information on productivity, and lender has to pay some cost to get this information. This will give the borrower some advantage, and make some possible transactions unable to happen. Townsend (1979) gives a micro model describing contract under costly state verification. Bernanke and Gertler (1989) and Carlstrom and Fuerst (1997) are two important works using this to explain business cycles (also called Financial Accelerator).

Add brief introduction for costly state verification literature.

7.3 Townsend, 79 JET

seminal paper in costly state verification, important, to be added.

7.4 Kiyotaki-Moore, 97 JPE

Kiyotaki, N. and J. Moore (1997). Credit cycles. *Journal of Political Economy* 105(2), 211–248

Another explanation for financial friction is collateral constraint. Important literatures include Kiyotaki and Moore (1997),

Kiyotaki and Moore (1997) has a simple idea. Suppose the borrower needs some collateral (say land) for borrowing. What is the effect of a temporary increase in land productivity? The direct effect is it will increase this period land price. More important is the indirect effect, the borrower now can buy more land since their collateral becomes more valuable. And thus in next period they can buy even more land. This long chain has a large effect on the initial price of the land. Thus financial sector can amplify the fluctuation of real economy.

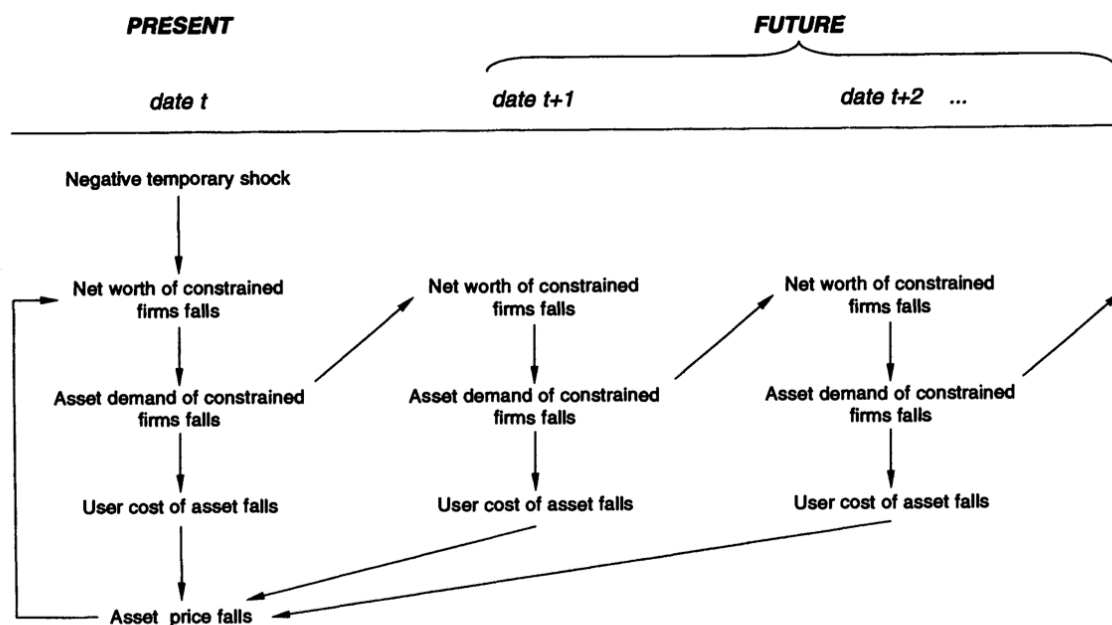


FIG. 1

Skills:

- How they deal with specific-targeted loan (i.e. the loan can only be used to purchase land)

7.5 Gorton and Ordonez, 14 AER

Gorton, G. and G. Ordoñez (2014, February). Collateral Crises. *American Economic Review* 104(2), 343–378

Motivation: How a small shock can sometimes have a very huge, sudden effect, while at other times the effect of the same size shock is small or nonexistent. Leverage per se is not enough to explain this. This paper explains how credit booms arise, leading to financial fragility where a small shock can lead to huge consequences. Thus, "tail risk" is endogenous. The same reason causes boom and huge recession, and the degree of recession is positively correlated with the boom lasting time.

Mechanism:

1. Firm needs land as collateral to borrow money from households. Land is either good or bad. Neither firm nor household know the quality of land, and checking fee is $\gamma > 0$. It can be shown that, when the probability of a land to be good is too high or too low, it will be optimal (for both firm and household) to not check the quality.
2. Then suppose there are idiosyncratic shocks to land. For each land in each period, there is λ prob to be unchanged, and $1 - \lambda$ to be changed. In case of changing, \hat{p} prob becoming good and $1 - \hat{p}$ becoming bad, independent of initial type. So we can take it as a \hat{p} type, it will become either 0 or 1 after checking. If \hat{p} lies in some range that no checking is appropriate, then uncertainty in the economy will accumulate. Notice such accumulation will lead to a boom.
3. Now think of an aggregate shock. All good type and \hat{p} type have weaker beliefs after such shock, changing to η and $\eta\hat{p}$ respectively. Notice that η and $\eta\hat{p}$ may have very different effect. It may be optimal to not check with η but optimal to check with $\eta\hat{p}$. Thus if an economy has accumulated a large portion of \hat{p} , i.e. a long boom, then many of them will choose to check, inducing a huge crisis.
4. Notice that it is not because of any *intrinsic* change in economy. The \hat{p} will change the real portion of good land in the economy only once. What really matters is that once the bad one becomes \hat{p} , then even it switches back to bad in following periods, it will still be taken as \hat{p} . That means, the \hat{p} type will accumulate in the economy.

Why this paper is important:

- Provide a way to endogenize information in general equilibrium model.

7.6 Williamson, 87 QJE

Williamson, S. D. (1987). Costly Monitoring, Loan Contracts, and Equilibrium Credit Rationing. *The Quarterly Journal of Economics* 102(1), 135–145

Williamson (1987) mainly wants to explain two things:

- Why is debt contract so popular?
 - By MM Theorem, we know that in a frictionless world, debt contract is equivalent to equity contract. But we notice that most fund is raised through debt.
 - Explain by moral hazard. Borrower has private information on outcome. And lender has to pay for checking.
- Why there is credit rationing?
 - If the borrowing demand exceeds lending supply, why doesn't the interest rate increase to clear the market?
 - Because checking is costly, so incentive compatibility requires interest rate not being too high. They prove that the optimal interest rate is hump shape.

An IC contract $\{S, K(w), R(w)\}$ contains three parts. S is the checking area. $K(w)$ is the transfer to lender when not checking. $R(w)$ is the transfer to lender when checking. Since the contract is IC, so the reported \hat{w} is just actual realized w .

There are three steps to prove the optimal contract is debt contract in this asymmetric information case.

1. Prove $K(w)$ equals a constant x . [easy]
2. Prove the IR constraint of lender will be tight. [easy]
3. Prove $R(w) = w$, i.e. lender will take away everything in checking case. [difficult]

In the third step, the idea is easy. Because if lender takes away everything in checking case, then the checking area can be smaller, thus can save checking cost. The difficulty is: when we lower x (to make checking area A smaller), the non-checking area B becomes larger. The changes are in two different directions and problem becomes subtle. **In discussion we use the total differentiation approach to solve this. Important.**

Techniques:

how to prove optimal contract format:

See above.

recursive structure in defining EQM:

One difficulty in determining the optimal contract is how to characterize outside options of lender. In equilibrium, When a lender decides whether to accept a contract from a specific borrower (entrepreneur), his outside option should be market interest rate r . Because entrepreneur can't identify lender, and all

entrepreneur are homogeneous ex ante, so if the expected return from a specific contract should be larger than mkt interest rate r . Thus the IR of lender is:

$$\int_A R(w)d(F(w)) + \int_B xd(F(w)) - \gamma Pr(A) \geq r$$

r has to be pinned down in the eqm. Because all agents have rational expectation, thus the market clearing r will be the same as their expectation.

The equilibrium is defined as:

•

7.7 Chari et.al, 14 AER

Chari, V. V., A. Shourideh, and A. Zetlin-Jones (2014, December). Reputation and Persistence of Adverse Selection in Secondary Loan Markets. *American Economic Review* 104(12), 4027–4070

Empirical Facts:

- simultaneous fluctuation between secondary loan market and underlying land market
- persistence of collapse in secondary loan market trade volume

Main Mechanism:

- adverse selection AND reputation
- key assumptions:
 - agent (bank) types persist over time
 - players are non-anonymous (possible to build reputation)
- adverse selection: can account for trade volume fluctuations correlated with underlying collateral value
- reputation:
 - without reputation concern, adverse selection will be solved quickly. Because the first stage features a separating equilibrium
 - with high enough reputation concern (patient enough), first stage cannot be fully separate, so adverse selection problem persists

7.8 Euro Crisis

This section is based on a course material of Wang Jue. Basically it discusses the theoretical concerns of Optimal Currency Area (OCA), why Europe adopts it, its success and its failure, and the recent Euro crisis.

Krugman, P. (2012). Revenge of the Optimum Currency Area. In *NBER Macroeconomics Annual 2012, Volume 27*, pp. 439–448. National Bureau of Economic Research, Inc

Krugman (2012) discusses the crisis of Euro area. Based on OCA theory by Mundell, there are two big things to look at: **labour mobility** and **fiscal integration**. Because areas are likely to suffer local idiosyncratic shocks.

- Without labour mobility, a state has to make wage fall to restore the full employment, this will be much easier with its own to devalue. But if labour mobility is high, then workers can just move to other states.
- Fiscal integration will also help. First because federal transfer is like a insurance, also because the borrowing cost for federal government is much lower.

The Euro crisis faces challenges in both parts.

- The labor mobility is not high enough.
- No fiscal integration. And the asymmetric shock is so severe, making fiscal burden so large that calls government solvency into question. (e.g. in Spain and Greek)
- The bank issue. There is no ‘last resort’, so the fear of sovereign default undermines confidence in private banks which hold these bonds. This is called *doom loops*.
 - Compare between euro area countries and pegged-to-euro countries. Austria and Finland uses euro, but Finland just pegged to euro. This flexibility makes interest rate much lower during the crisis.

Baldwin, R. E. and F. Giavazzi (2015). *The Eurozone Crisis: A Consensus View of the Causes and a Few Possible Remedies*. CEPR Press

7.9 Seminar Notes

Cryptocurrency

- By Jesús Fernández-Villaverde and Daniel Sanches 2017 .Use Lagos and Wright (2005) framework.
 - <http://economics.sas.upenn.edu/~jesusfv/Cryptocurrencies.pdf>

- Not positive on private currencies. First, competitive private money will not provide price stability, either with a profit-maximizing entrepreneurs or through automaton.
- Even with price stability, value of private moneys may go zero in infinite equilibriums.
- Purely private money system does not provide social optimal quantity of money.

Chapter 8

Methodology

Reductionism, and Micro Foundations of Economics:

Economists are fascinated in forming ‘micro foundations’ for almost all models. But do we really need this, does this effort pay?

As pointed in famous paper Anderson (1972): ‘*The ability to reduce everything to simple fundamental laws doesn’t imply the ability to start from those laws and reconstruct the universe... At each level of complexity entirely new properties appear... There is no useful parallel to be drawn between the way in which complexity appears in simplest cases of many modies and chemistry and culture ones.*’ The aggregated economy is so complicated that may not be reduced to individual decision problems.¹

Another problem worth thinking is the relationship between Sonnethein-Debreu theorem and macro economics models.

Meaning of Decision Theory:

(This part is heavily influenced by Henrique’s class.)

Why we need to do decision theory? Take an *inverse engineering* perspective. Suppose we have a model, which uses utility function to describe people’s decision choices. That is, $u : X \rightarrow \mathbb{R}$. And we define $C(B) = \arg \min_{x \in B} u(x) = \{x \in B | u(x) \geq u(y), \forall y \in B\}$. We can map such functions to a *equivalent class*, and can further map it to a *preference relation* (a subset of relations).

The problem is such mapping is not surjective. Every utility function has a corresponding equivalence and a preference relation, but there is some preference relation that cannot be represented by a utility function. Thus we need to put some restrictions over preference relation.

The first relation is that the cardinality of indifference curves cannot be larger than \mathbb{R} . This is easy to understand.

But this restriction is not enough. Think of lexicographic preference, every point in the two-dimension

¹Zhao Dingxing has an article discussing this problem.
See <https://www.evernote.com/l/ATU0IdQb8hBM5r3QQv8cTK52Amq4-q1wb0E>

space is a indifference class. The cardinality of which is R^2 , equivalent to cardinality of \mathbb{R} . But we know that such preference do not have a corresponding utility function. We can further restrict preference to *order separable* relations, then by Debreu Thm, we can show the *iff* between utility and preference relations.

Definition 8.0.1 (Order Separable). A preference relation defined over X is order separable if $\exists Y \subseteq X$ countable s.t. $\forall x_1, x_3 \subseteq X$, we have $x_1 \succ x_2 \Rightarrow \exists y \in Y$ s.t. $x_1 \succ y \succ x_2$.

Important here is that by assuming the range of utility function is \mathbb{R} (such a weak condition), we actually put some restrictions over preference.

Calibration:

Why use calibration instead of estimation?

- Because we live in a ‘tenth-best’ world, and our model only catches a small part of the world. Use calibration we can fit the moments that our model fail to catch, in this sense to be more accurate.

May use calibration parameters as an intuitive check of whether the code we wrote is correct.

Research Reproducibility

Maniadis and Tufano (2017) claims two aspects of reproducibility problem. First, it can be seen as an ‘evidence game’, where researchers have some private information and evaluators have to design a incentive compatible mechanism for publishing and granting. Second, empirical methods from meta-research may help. (I don’t understand the second part.)

Maniadis et al. (2017) in the same issue provides more detailed guide on what will affect the credibility of scientific research. It raises a concept post-study probability,

$$PSP = \frac{(1 - \beta)\pi}{(1 - \beta)\pi + \alpha(1 - \pi)}$$

where α is significance level, $1 - \beta$ is power, and π is the probability that a research question is actually true. π can be think of as *prior* given by theory. This is the fraction of ‘truly true and reported true’ divided by ‘all reported true’. Easy to see, when π is low, the PSP can be very low even with low α . This indicates that with those fields where theory can provide little guide, we have to be very skeptical to the ‘rejection’ outcomes. They are very likely to be ‘false positive’.

Part II

Mathematics

Chapter 9

General (mathematical) Skills

9.1 Log-Linearization

Motivation

Why we need log linearizations? Because for most nonlinear discrete dynamic programming problems, we fail to find a closed solution. Thus we have to use numerical method or to find a approximation. By using log-linearization, we transform the nonlinear problem to a linear problem (around the steady state), and we know how to solve the linear difference equations.

Another advantage of log-linearization is, the new variables are in forms of $\frac{x-x^*}{x^*}$, which is interpretable.

9.1.1 How to do Log-Linearization

Log linearization is a common method to approximate non-linear function using Taylor expansion.

When we do linearization in macroeconomics, one key is to find **steady state** of the model. Then we do linearization around the steady state.

The basic idea is, for functions like:

$$f(x) = \frac{g(x)}{h(x)}$$

take log in both sides:

$$\ln(f(x)) = \ln(g(x)) - \ln(h(x))$$

By Taylor expansion, for a smooth function $f(x)$ we have:

$$f(x) = f(x^*) + \frac{f'(x^*)}{1!}(x - x^*) + o(1)$$

Thus

$$\ln f(x) \approx \ln f(x^*) + \frac{f'(x^*)}{f(x^*)}(x - x^*)$$

The above follows from the fact that $\frac{d \ln(f(x))}{dx} = \frac{f'(x)}{f(x)}$. Thus,

$$\ln f(x) - \ln f(x^*) = \frac{f'(x^*)}{f(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*}$$

And the equation becomes:

$$\ln f(x^*) + \frac{f'(x^*)}{f(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*} = \ln g(x^*) - \ln h(x^*) \frac{g'(x^*)}{g(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*} - \frac{h'(x^*)}{h(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*}$$

$$\frac{f'(x^*)}{f(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*} = \frac{g'(x^*)}{g(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*} - \frac{h'(x^*)}{h(x^*)} \cdot x^* \cdot \frac{x - x^*}{x^*}$$

Dedine $\hat{x} = \frac{x - x^*}{x^*}$, and we are done.

$$\frac{f'(x^*)}{f(x^*)} x^* \cdot \hat{x} = \frac{g'(x^*)}{g(x^*)} x^* \cdot \hat{x} - \frac{h'(x^*)}{h(x^*)} x^* \cdot \hat{x}$$

Notice, when we use \hat{x} in place of $x - x^*$ in the approximation, don't forget to times x^* !!

9.1.2 log-linearization when there is expectation (An Example)

It may be annoying to do log-linearization when there is expectation, because in general, log and expectation cannot switch. But if we don't switch, we cannot do the log-linearization as previous. For example, see Adda&Cooper(2003) CH5, page 115.

Suppose we have the following Euler equation:

$$u'(c) = \beta E_{A'|A} u'(c) [A f'(k') + (1 - \delta)]$$

where $c = A f(k) + (1 - \delta)k - k'$.

These two expressions, along with the evolution of A, defines a system of equations.

In this system, actually we do not have to take log, but directly do Taylor expansion at c^* , x^* and k^* .

We get:

$$\begin{aligned} u'(c^*) + u''(c^*) c^* \hat{c}_t &= \beta E_{A'|A} [u'(c^*) (\bar{A} f'(k^*) + (1 - \delta)) + (\bar{A} f'(k^*) + (1 - \delta)) u''(c^*) c^* \hat{c}_{t+1} \\ &\quad + u'(c^*) f'(k^*) \bar{A} \hat{A}_{t+1} + u'(c^*) \bar{A} f''(k^*) k^* \hat{k}_{t+1}] \quad (9.1) \end{aligned}$$

The first term in LHS and the first term in RHS cancel each other. And we divide both sides by $u'(c^*)$, and we will get:

$$\frac{u''(c^*)c^*}{u'(c^*)}\hat{c}_t = \beta E_{A|A'}\left[\frac{u''(c^*)c^*}{u'(c^*)}\hat{c}_{t+1}\frac{1}{\beta} + f'(k^*)\bar{A}\hat{A}_t + 1 + f''(k^*)k^*\hat{k}_{t+1}\right] \quad (9.2)$$

Notice, we derive this based on the fact $\frac{1}{\beta} = \bar{A}f'(k^*) + (1 - \delta)$. Because by FOC

$$u'(c) = \beta E_{A'|A} V'(k')$$

To derive $V'(k')$, we first derive $V'(k)$, by Envelop THM we get:

$$V'(k) = u'(c)[Af'(k) + (1 - \delta)]$$

Substitute this into FOC we get:

$$u'(c) = \beta E_{A|A'}[u'(c')(A'f'(k') + (1 - \delta))]$$

In this approach we fix A at the mean \bar{A} , and thus the steady state satisfy:

$$1 = \beta[\bar{A}f'(k^*) + (1 - \delta)]$$

Here, the introduce of \bar{A} is to describe the steady state in existence of shock. We can see \bar{A} as the "long term" growth of the economy, and the we investigate what's the optimal steady k^* under this long-term growth rate.

But I still don't know why we can cancel the expectation in the RHS.

The equation 9.2 is the linear function we want.

Some Comments on the above method:

Although it seems that the above method uses no log-linearization, actually it does!

(to be added later)

9.2 Supermodularity

Reference:

- Sobel's Notes: <http://econweb.ucsd.edu/~jsobel/205f10/notes12.pdf>
- A more detailed note: <https://pages.wustl.edu/files/pages/imce/nachbar/monotonecomparativestatics.pdf>

- Topkis, D. (1998). *Supermodularity and Complementarity*, 1998. Princeton University Press

Chapter 10

Probability

10.1 Bayesian Statistics

10.1.1 Bayesian Updating

Theorem 10.1.1 (Bayesian Update in Odds). *For hypothesis H , data d and prior odds $O(H)$. We have posterior odds $O(H|d) = O(H) \cdot \frac{Pr(d|H)}{Pr(d|H^c)}$, where the second term is called Bayes Factor.*

Proof:

$$\begin{aligned} O(H|d) &= \frac{Pr(H|d)}{Pr(H^c|d)} \\ &= \frac{Pr(d|H)Pr(H)}{Pr(d|H^c)Pr(H^c)} \\ &= \frac{Pr(d|H)}{Pr(d|H^c)} \cdot O(H) \end{aligned}$$

□

Chapter 11

Numerical Methods

Computation language:

- http://economics.sas.upenn.edu/~jesusfv/comparison_languages.pdf
 - discusses efficiency of different languages
 - hybrid language can significantly increase efficiency

Matlab Speed Up:

- Microeconometrics and Matlab: an Introduction, Ch-10
- vectorize and preallocate values
 - Cooper's growth model uses clever vectorizing, avoid one loop
- profiling code: profile on, main, profile off
- parallel computing

11.1 Numerical Maximization

Newton-Raphson Method:

Idea: use quadratic function to approximating real function at current guess, then find the maximization point of this quadratic function as next guess.

$$\beta_{t+1} = \beta_t + \lambda(-H_t^{-1})g_t$$

where H_t is the Hessian of the function to be evaluated at β_t , and g_t is the Jacobian at β_t , λ is the step size, generally may take 1.

BHHH Method:

Idea: almost the same as Newton-Raphson, but use the outer product to approximate Hessian.

$$\beta_{t+1} = \beta_t + \lambda(-B_t^{-1})g_t$$

where $B_t = \frac{1}{N} \sum_{n=1}^N s_n(\beta_t)s_n(\beta_t)'$ and $s_n(\beta_t) = \partial P_n(\beta)/\partial \beta$ evaluated at β_t , and $g_t = \frac{1}{N} \sum_{n=1}^N s_n(\beta_t)$.

Key is, when evaluated at maximum value, s_n should be zero. Thus B_t evaluated at maximum is the variance of s_n in the sample, which is just the Hessian of the sample likelihood function. This is just *information equality*.

pros: BHHH is generally faster than Newton-Raphson near the maximum. Because s_n have to be calculated anyway, but in NR procedure we have to calculate the second derivatives as well. *But seems this is not the case, in experiments BHHH always slower than Newton-Raphson.*

cons: BHHH is generally slower when β_t is far from true value. Because at these points B_t is not a good approximate of H_t .

- Root Solving
- Integration
 - Gaussian quadrature
 - Sparse grid
 - Monte Carlo integration
 - * antithetic accelerating
 - * importance sampling
 - Quasi Monte Carlo

Chapter 12

Miscellany

Presentation:

- Always motivate very clearly, both from economics facts and from literature.
 - In macro, a graph generally helps
- If possible, always use a simple model at the beginning
 - To pin down the concepts/definitions
 - To show the mechanism intuitively
- Understand the result both mathematically and intuitively
- When writing, write horizontally!

Bibliography

- Akerberg, D. A., K. Caves, and G. Frazer (2015). Identification Properties of Recent Production Function Estimators. *Econometrica* 83(6), 2411–2451.
- Anderson, P. W. (1972). More is different. *Science* 177(4047), 393–396.
- Baldwin, R. E. and F. Giavazzi (2015). *The Eurozone Crisis: A Consensus View of the Causes and a Few Possible Remedies*. CEPR Press.
- Baye, M. R., B. De los Santos, and M. R. Wildenbeest (2013). The Evolution of Product Search. *The Journal of Law, Economics Policy* 9(2), 1–21.
- Ben-Porath, E., E. Dekel, and B. L. Lipman (2014, December). Optimal Allocation with Costly Verification. *American Economic Review* 104(12), 3779–3813.
- Bernanke, B. and M. Gertler (1989, March). Agency Costs, Net Worth, and Business Fluctuations. *American Economic Review* 79(1), 14–31.
- Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica* 63(4), 841–890.
- Berry, S. T. (1994, July). Estimating Discrete-Choice Models of Product Differentiation. 25(2), 242–262.
- Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther (2010, December). The financial reporting environment Review of the recent literature. *Journal of Accounting and Economics* 50(2-3), 296–343.
- Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy* 100(5), 992–1026.
- Carlstrom, C. T. and T. S. Fuerst (1997, December). Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis. *American Economic Review* 87(5), 893–910.
- Chari, V. V., A. Shourideh, and A. Zetlin-Jones (2014, December). Reputation and Persistence of Adverse Selection in Secondary Loan Markets. *American Economic Review* 104(12), 4027–4070.

- Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2017). Revealed price preference: Theory and stochastic testing.
- Dekel, E. and B. L. Lipman (2010). How (not) to do decision theory. *Annual Review of Economics* 2(1), 257–282.
- Dye, R. A. (1985). Disclosure of Nonproprietary Information. *Journal of accounting research* 23(1), 123–145.
- Erdem, T. and M. P. Keane (1996). Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science* 15(1), 1–20.
- Fudenberg, D. and D. K. Levine (1993). Self-Confirming Equilibrium. *Econometrica* 61(3), 523–545.
- Gale, D. and L. S. Shapley (1962, January). College Admissions and the Stability of Marriage. *The American Mathematical Monthly* 69(1), 9–15.
- Gandhi, A., S. Navarro, and D. Rivers (2017). On the Identification of Gross Output Production Functions.
- Gentzkow, M., J. Shapiro, and M. Taddy (2017, May). Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech.
- Gentzkow, M., M. Taddy, and B. T. Kelly (2017, February). Text as Data. pp. 1–53.
- Glazer, J. and A. Rubinstein (2004). On Optimal Rules of Persuasion. *Econometrica* 72(6), 1715–1736.
- Gorton, G. and G. Ordoñez (2014, February). Collateral Crises. *American Economic Review* 104(2), 343–378.
- Green, E. J. and R. H. Porter (1984). Noncooperative Collusion under Imperfect Price Information. *Econometrica* 52(1), 87–100.
- Green, J. R. and J.-J. Laffont (1986). Partially Verifiable Information and Mechanism Design. *The Review of Economic Studies* 53(3), 447.
- Grimmer, J. and B. M. Stewart (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21(3), 267–297.
- Grossman, S. J. (1981, December). The Informational Role of Warranties and Private Disclosure about Product Quality. *The Journal of Law & Economics* 24(3), 461–483.
- Hart, S., I. Kremer, and M. Perry (2017, March). Evidence Games: Truth and Commitment. *American Economic Review* 107(3), 690–713.
- Hausman, J., G. Leonard, and J. D. Zona (1994, April). Competitive Analysis with Differentiated Products. *Annales d'Économie et de Statistique* (34), 159–180.

- Keane, M. P. and K. I. Wolpin (1994). The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence. *The Review of Economics and Statistics* 76(4), 648–672.
- Kiyotaki, N. (2011). A mechanism design approach to financial frictions. *manuscript*.
- Kiyotaki, N. and J. Moore (1997). Credit cycles. *Journal of Political Economy* 105(2), 211–248.
- Kremer, I., Y. Mansour, and M. Perry (2014). Implementing the “Wisdom of the Crowd”. *Journal of Political Economy* 122(5), 988–1012.
- Krugman, P. (2012). Revenge of the Optimum Currency Area. In *NBER Macroeconomics Annual 2012, Volume 27*, pp. 439–448. National Bureau of Economic Research, Inc.
- Levinsohn, J. and A. Petrin (2003, April). Estimating Production Functions Using Inputs to Control for Unobservables. *The Review of Economic Studies* 70(2), 317–341.
- Lewbel, A. (2016). The Identification Zoo-Meanings of Identification in Econometrics. *Forthcoming on Journal of Economic Literature*.
- Low, H. and C. Meghir (2017, May). The Use of Structural Models in Econometrics. *Journal of Economic Perspectives* 31(2), 33–58.
- Maniadis, Z. and F. Tufano (2017, October). The Research Reproducibility Crisis and Economics of Science. *The Economic Journal* 127(605), F200–F208.
- Maniadis, Z., F. Tufano, and J. A. List (2017, October). To Replicate or Not To Replicate? Exploring Reproducibility in Economics through the Lens of a Model and a Pilot Study. *The Economic Journal* 127(605), F209–F235.
- Milgrom, P. R. (1981, October). Good News and Bad News: Representation Theorems and Applications. *The Bell Journal of Economics* 12(2), 380–391.
- Olley, G. S. and A. Pakes (1996, November). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica* 64(6), 1263–1297.
- Quan, T. W. and K. R. Williams (2017). Product Variety, Across-Market Demand Heterogeneity, and the Value of Online Retail. *SSRN Electronic Journal*, 1–74.
- Quigley, D. and A. Walther (2017, May). Inside and Outside Information.
- Rey, P. and J. Stiglitz (1995, October). The Role of Exclusive Territories in Producers’ Competition. *The RAND Journal of Economics* 26(3), 431–451.

- Schmalensee, D. S. E. and Richard (2013, October). The Antitrust Analysis of Multi-Sided Platform Businesses. *manuscript*, 1–45.
- Shum, M. (2016). *Econometric Models for Industrial Organization*. World Scientific.
- Smith, L. and P. Sørensen (2000). Pathological outcomes of observational learning. *Econometrica*.
- Topkis, D. (1998). *Supermodularity and Complementarity, 1998*. Princeton University Press.
- Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory* 21(2), 265–293.
- Williamson, S. D. (1987). Costly Monitoring, Loan Contracts, and Equilibrium Credit Rationing. *The Quarterly Journal of Economics* 102(1), 135–145.